Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Caroline Plott April 15, 2015

Investigating the Role of *spr-5* and *met-2* in Exogenous Gene Expression

By

Caroline Plott

Dr. David Katz
Adviser

Biology Department

Dr. David Katz

Adviser

Dr. William Kelly

Committee Member

Dr. Tracy McGill

Committee Member

Dr. Rachelle Spell

Committee Member

2015

Investigating the Role of *spr-5* and *met-2* in Exogenous Gene Expression

By

Caroline Plott

Dr. David Katz

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Biology Department

2015

Abstract

Investigating the Role of *spr-5* and *met-2* in Exogenous Gene Expression

By Caroline Plott

Both *spr-5* and *met-2* are implicated in the epigenetic silencing of endogenous, germline-specific genes in *Caenorhabditis elegans* (*C. elegans*) by their demethylase and methyltransferase activity, respectively. To elucidate the role that these genes serve in silencing of exogenous genes in the *C. elegans*, we asked if *gfp* (green fluorescent protein) transgenes escape germline silencing in *spr-5* (suppressor of presenilin) and *spr-5;met-2* (histone methyl transferase-like) mutants. Surprisingly, we found that when we crossed an extrachromosomal array containing a *let-858::gfp* transgene into a homozygous *spr-5;met-2* mutant background not only was there no expression of this transgene in the germline, the sterile double mutant progeny had dramatically reduced expression of LET-858::GFP in somatic cells, suggesting somatic silencing of the transgene construct. This result indicates that the extrachromosomal array was silenced somatically as well as in the germline. Furthermore, out of the approximately 230 *spr-5* mutants injected with a *gfp::cdk-1* (cyclin dependent kinase) transgene, only five *C. elegans* lines were produced that inheritably expressed the transgene construct somatically, further implicating SPR-5 in active somatic expression. These findings implicate *spr-5* and *met-2* in an indirect role in maintaining transgenes in an active state in the soma. These findings may also indicate that *spr-5* and *met-2* play a role in maintaining the distinction between soma and germline cell identity.

Investigating the Role of *spr-5* and *met-2* in Exogenous Gene Expression

By

Caroline Plott

Dr. David Katz
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Biology Department

2015

Acknowledgements

I am grateful to Dr. David Katz for training me, helping me to design experiments, providing comments, and being a wonderful mentor. Thank you also to Dr. Teresa Lee for her help with the mating experiments, comments, and for her constant moral support. Thank you to Dr. Karen Schmeichel for mentoring me and creating the *spr-5* mosSCI strain. Thank you to Dr. Kelly, Dr. McGill and Dr. Spell for invaluable input and suggestions. Thank you to John Jeong Ahn for providing advice about the injection procedure. Thank you to the Kelly lab for letting us use their microinjection equipment. Thank you also to the Mello and Baumeister labs for providing *C. elegans* strains and plasmids. Thank you to the Benian lab for providing the EG6699 strain *C. elegans*. Strains were also obtained from the *Caenorhabditis* Genetic Center.

I am also grateful to my family and friends who supported me through this journey.

Table of Contents

Figures

Tables

Background

Epigenetic Identity and Transcriptional Gene Regulation

Organisms must dynamically adjust gene transcription in order to meet developmental needs, changing environmental conditions, and to regulate non-self DNA. One of the ways in which this transcriptional regulation is achieved is through the use of epigenetic modifications. Epigenetic regulation is distinct from other types of genetic regulation because it functions through the physical remodeling of the genetic material around histone proteins. These chemical changes can control gene expression levels by altering the accessibility of the gene to protein factors (Kouzarides, 2007). Every eukaryotic cell contains negatively charged DNA in its nucleus that is associated with nucleosomes. Nucleosomes are positively charged protein octamers containing two sets of each type of histone protein: H2A, H2B, H3, and H4 (Kornberg, 1974). These histones have N-terminus tails that can be modified by a host of chemical alterations, including acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ADP ribosylation, deamination, and proline isomerization (Kouzarides, 2007).

We were specifically interested in the effects of the loss of two epigenetic modifier proteins, SPR-5 and MET-2. SPR-5 is a demethylase that removes methyl groups from histone 3 lysine residue 4 (H3K4) (Shi et al., 2004). MET-2 is a methyltransferase, adding methyl groups to histone 3 lysine residue 9 (H3K9) (Bessler et al., 2010). Methylation is a particularly interesting method of gene regulation because it is reversible and potentially dynamic. Cells can mono-, di- and tri- methylate both lysine and arginine residues on histone tails. Therefore, methylation may serve an important role in the regulation of self and non-self DNA as well as in the distinction between somatic and germline cell identities (Kouzarides, 2007). We are also interested in the role that these genes play in the regulation of somatic genes. Previously, the

Katz lab has done extensive work to elucidate the role of *spr-5* and *met-2* in germline specific gene regulation (this work which will be discussed later).

Within the realm of epigenetic regulation, an additional distinction must be made between regulation of self and non-self DNA. Self, or endogenous, DNA can be defined as genes that are found in the vast majority of the population, that have been present in that species for a long span of time, or that are present in the organism's published reference genome. Transposons are considered to be endogenous DNA because they fit both of these criteria. Conversely, exogenous DNA is that which is not found in the majority of the population. Exogenous DNA can originate from the genome of another organism, and can be introduced by a vector such as a virus, or experimentally.

Previous work on the epigenetic regulation of self-DNA and of non-self DNA has indicated that the mechanism by which these two processes are carried out is related. Organisms' ability to regulate the silencing and expression of transposons and other self-DNA elements appears to be affected when genes which contribute to non-self DNA regulation are mutated (Shirayama et al., 2012)

In addition to the distinction between endogenous and exogenous DNA, there is a distinction between the methods of regulation in somatic and germline cells. Germline cells include sperm and egg cells, which fuse to create an undifferentiated oocyte. Effectively, all of the other cells, from epithelial cells to neurons, in an organism are somatic. Both germline and somatic cells must be able to distinguish between and regulate endogenous and exogenous DNA. However, the challenges faced by these two cell types are often quite different. In terms of endogenous gene regulation, somatic cells must constantly regulate transcriptional patterns to respond to environmental challenges. For example, a response to environmental changes that has

been well studied in multiple model systems is the epigenetic and transcriptional changes associated with the exposure to high heat, called heat shock. In this case, chaperone proteins, along with other necessary genes, must be quickly activated in order to avoid cell death (Hightower, 1991).

Cells in the germline, however, face an even more daunting type of endogenous gene regulation. Before fertilization, the highly differentiated sperm and egg cells each have a specific gene expression pattern. Upon fertilization, gene expression must be dramatically altered to create an undifferentiated diploid zygote cell. Furthermore, effective regulation of DNA is particularly important in the germline, as incorrect expression of transposable genes and non-self DNA can lead to the spread of harmful elements in the genome of the next generation, reducing fitness. Therefore, cells in the germline are under higher selection to properly silence any elements in the germline that could be detrimental to the next generation (Katz et al., 2009; Shirayama et al., 2012).

piRNA Mediated Epigenetic Regulation

There are three small non-coding RNAs involved in gene regulation: microRNA (miRNA), small interfering RNA (siRNA) and Piwi-acting RNA (piRNA). miRNA and siRNA predominately act on the post-transcriptional level by targeting mRNA transcripts for degradation (Kutter and Svoboda, 2008). Both miRNAs and siRNAs are cut into smaller pieces approximately 21-22 nucleotides long by the Dicer protein (Bartel, 2004; Kutter and Svoboda, 2008). The processed short dsRNA sequences are made single stranded, and loaded onto an Argonaut protein which is part of the larger RNA-induced silencing complex (RISC). The

complementary base pairing ability of the siRNA or miRNA guides the RISC complex to its target (Bartel, 2004).

miRNAs and siRNAs differ in their biogenesis and targets. miRNA is directly transcribed from non-coding DNA in the nucleus, whereas siRNA is created using mRNA templates (Bartel, 2004). Furthermore, miRNAs are involved mainly in the regulation of endogenous gene transcripts, notably during development (Bartel, 2004; Kutter and Svoboda, 2008). siRNAs, on the other hand, are involved in the regulation of both endogenous and exogenous transcripts, and are produced from long dsRNA (Bartel, 2004; Shirayama et al., 2012). Both miRNA and siRNA have both been implicated in silencing regulation on the transcriptional level (Morris et al., 2004).

miRNA and siRNA, and piRNA all carry out their silencing functions by guiding AGO proteins and the rest of the silencing or degradation protein complex to the final target transcript or chromatin location (Holoch and Moazed, 2015). However, unlike siRNAs and miRNAs, piRNAs are implicated in epigenetic regulation, are initially single stranded, and are not processed by Dicer (Holoch and Moazed, 2015; Kutter and Svoboda, 2008). In the siRNA pathway, the RDE-1 protein brings RdRP to the mRNA template to make the 22G targeting RNA. However the RDE-1 protein is required neither for transposon nor exogenous gene silencing, suggesting that the silencing mechanism of these elements is distinct from the miRNA and siRNA pathway (Shirayama et al., 2012). Work by the Mello lab has implicated the piRNA pathway in the initial silencing of non-self DNA (Shirayama et al., 2012).

Transgenes can be present in an organism either as a multi-copy exogenous array or integrated into the genome. Initial work in the field of exogenous gene silencing predominately focused on the silencing of multi-copy transgenic arrays. While *C. elegans* can transmit

transgenic arrays to the next generation, *C. elegans* generally silence multi-copy transgenic arrays in the germline (Kelly et al., 1997).

In order to circumvent the silencing of multi-copy transgenic arrays, a method was developed by the Jorgensen lab to introduce single copy genes into *C. elegans*. Unlike multi-copy arrays, most single copy insertions are not silenced in the germline, suggesting that the non-self gene silencing mechanism may depend on copy number. Nonetheless, some single copy transgenes are still silenced in the germline. The mechanism for silencing of multi-copy arrays and integrated genes is thought to be related. When a single copy construct consisting of an endogenous gene fused to an exogenous gene, *cdk-1::gfp*, was integrated into *C. elegans*, the Mello lab found that the exogenous gene was silenced in the germline, while the endogenous gene was unaffected. Conversely, when multi-copy arrays consisting of an exogenous and endogenous gene are introduced into *C. elegans*, both the exogenous and endogenous genes are silenced or down-regulated (Shirayama et al., 2012).

met-2 *in Epigenetic Germline Regulation*

The ability to affect chromatin structure through histone modifications provides the cell with a dynamic method of regulating genes, as well as passing on, or not passing on, epigenetic information to the next generation. One such histone modification that can both silence genes in actively transcribed regions as well as establish a larger more global repressive chromatin state is H3K9 methylation (Kouzarides, 2007). In *C. elegans,* some H3K9 dimethylation is carried out by the methyltransferase MET-2 (Bessler et al., 2010). MET-2 is a *C. elegans* analog of human SETDB1 which is in the SET-domain protein superfamily (Poulin et al., 2005). The SETDB1 protein, like many proteins that catalyze H3K9 methylation, has a SET protein domain, which is

coded for by an mRNA sequence that is highly conserved in most eukaryotic organisms (Dillon et al., 2005). Epigenetically, the transfer of a methyl group onto H3K9 results leads to a repressed state (Kouzarides, 2007). Interestingly, when H3K9 methylation is placed on the coding region, it has also been shown to lead to a more active chromatin environment (Vakoc et al., 2005).

In *C. elegans*, *met-2* is located on Chromosome III. *met-2* mutants have smaller brood sizes than wild-type *C. elegans*, display an increasing sterility phenotype, and after 28 generations have a severely reduced brood size (Kerr et al., 2014). *met-2* animals have decreased levels of H3K9 methylation on genes that are targeted by *spr-5*, suggesting that these two proteins work in conjunction (Kerr et al., 2014). *met-2* has only been shown to be expressed in the germline, however it is also implicated in the development pattern of somatic vulval cells, indicating that it has a role in somatic cells (Anderson and Horvitz, 2007; Kerr et al., 2014).

spr-5 *in Epigenetic Germline Regulation*

In order to regulate H3K4 dimethylation, cells have developed mechanisms to demethylate histone residues through the use of lysine demethylases (Kouzarides, 2007). Some histone methylation marks are readily reversible and modifiable demethylation through the action of histone demethylase enzymes (Shi et al., 2004). Such reversible, enzyme catalyzed histone methylation is associated with transcriptional regulation. However, methylation in DNA regions where more long-term silencing is necessary is not readily reversible by such enzymatic processes, and demethylation in these regions is carried out through removal and replacement of the entire histone (Bannister et al, 2002).

Shi et al. identified the first H3K4 demethylase enzyme, *lsd-1* (lysine specific demethylase), in 2004. LSD1 demethylases mono- and di-methylated lysine residues (Shi et al., 2004). Demethylation is carried out by proteins related to the amine oxidase family through oxidation of the amine group (Shi et al., 2004).

LSD1/KDMI is one such demethylase in humans, and its activity results in an increasingly repressed chromatin environment (Shi et al., 2004). In addition to its demethylase activity, *lsd1* has been implicated in meiotic processes in both chromatin remodeling as well as in the repair of double strand DNA breakages (Nottke et al., 2011). The *C. elegans lsd1* homolog is *spr-5* (Katz et al., 2009). In *C. elegans, spr-5* mRNA is thought to be deposited maternally in the fertilized zygote (Katz et al., 2009).

LSD1/KDMI associates with a six-protein complex called the co-REST-HDAC that co-represses with the RE1 silencing transcription factor (REST) complex (Lakowski et al., 2006, Shi et al., 2004). The coREST-HDAC complex is associated with repression of transcription, and has been found to regulate genes implicated in neuronal development in multiple model systems (Lakowski et al., 2006). REST is responsible for assisting in assembling the BHC complex, a histone deacetylase complex which down regulates the expression of neuronal genes on the transcriptional level in non-neuronal cells (Lee et al., 2005; Shi et al., 2004). The roles of REST, co-REST, and LSD1 in epigenetic regulation are not fully understood.

The LSD1 protein sequence is very similar to that of proteins in the family of flavin adenine dinucleotide (FAD) amine oxidases, suggesting that its chemical mechanism of demethylation is similar to the chemical mechanism of amine oxidases (Shi et al., 2004). The LSD1 demethylase reaction occurs by breaking the alpha nitrogen carbon bond via an oxidation

reaction resulting in an imine group. The imine is then converted to an amine by hydrolysis, also producing formaldehyde (Shi et al. 2004).

spr-5 *and* met-2 *in Epigenetic Exogenous Germline Regulation*

In addition to the regulating the expression of endogenous genes, organisms must properly silence exogenous sequences. Previous work has indicated that there may be overlap between the epigenetic mechanism used by organisms to silence endogenous and exogenous sequences (Ketting et al., 1999). Shirayama et al. recently described one such kind of epigenetic regulation that regulates both endogenous genes as well as transposable elements as "RNA-induced epigenetic silencing" (RNAe) (Shirayama et al., 2012). RNAe is established through the piRNA pathway, and is unique in that it leads to stable, heritable gene silencing (Shirayama et al., 2012).

Historically, it has been difficult to study the mechanism by which *C. elegans* silences exogenous genes in the germline because the common method of introducing exogenous genes was through the generation of multi-copy exogenous arrays, which are almost always silenced in the *C. elegans* germline (Kelly et al., 1998). These silenced transgene constructs have increased levels of H3K9me, a histone modification implicated in silencing (Kelly et al., 2002). The Mos1-mediated Single Copy Insertion (mosSCI) method allows for the insertion of a single copy of the gene of interest (Frøkjær-Jensen et al., 2008). This method results in germline expression of an integrated transgene when a relatively small exogenous sequence, such as those containing *flag* (which produces an eight amino acid protein product), is used (Shirayama et al., 2012; Kaltwasser et al., 2002). However, larger transgene sequences, such as those containing *gfp*

(which produces a 238 amino acid protein product), are often silenced even when only present as a single copy (Chalfie et al., 1994; Shirayama et al., 2012).

piRNAs have been implicated in germline transposon silencing (Lee et al., 2012). In their model, Shirayama et al. proposed that piRNAs are loaded onto Piwi Argonaut proteins that target the transcript to be silenced (Shirayama et al., 2012). There are two Piwi Argonaut proteins in *C. elegans*, PRG-1 and PRG-2, of which only PRG-1 is understood to have a significant role in germline silencing (Das et al., 2008). Shirayama et al. showed that *prg-1* mutants displayed germline de-silencing of integrated single copies of constructs composed of a large exogenous gene (*gfp*) annealed to an endogenous gene (either *cdk-1* or *csr-1*). This result implicated *prg-1* in the initiation of germline silencing and the RNAe pathway (Shirayama et al., 2012).

Both *spr-5* and *met-2* are expressed in the *C. elegans* germline, and *spr-5* and *met-2* mutants become progressively sterile over multiple generations. This increase in sterility is associated with an accumulation of H3K4 methylation in the germline located on genes involved in spermatogenesis. Furthermore, *met-2* adds H3K9 methylation to some of the same spermatogenesis genes from which *spr-5* removes H3K4 methylation (Kerr et al., 2014). Additionally, *spr-5;met-2* double mutants have an aggravated sterility phenotype, becoming completely sterile after only two generations, and have increased accumulation of H3K4 methylation (Kerr et al., 2014). It was demonstrated that this H3K4me2 accumulation led to the improper transcription of endogenous spermatogenesis genes in the germline (Kerr et al., 2014). These observations suggest that SPR-5 and MET-2 are involved in the same pathway (Kerr et al., 2014).

In addition to their work on endogenous gene regulation, Katz et al. showed that *spr-5* mutants de-silence a multi-copy H2A::GFP transgene in the germline (Katz et al., 2009). This

result was surprising since wild-type *C. elegans* usually silence multi-copy transgenes in the germline. This observation suggested that *spr-5* was also involved in the mechanism for silencing exogenous gene elements (Katz et al., 2009).

Green Fluorescent Protein Marker

Green fluorescent protein (GFP) was initially found to be present in *Aequorea victoria (A. victoria)*, a species of jellyfish. GFP is a relatively large protein (238 amino acids) (Chalfie et al., 1994). In vivo, *A. victoria* produces energy in the form of blue light when calcium ions bind to the aequorin protein. The energy released from this interaction excites GFP, which consequently produces green light. GFP also produces green light when the exciting energy is derived from an externally applied blue light source such as ultraviolet light. This unique light producing property of the GFP protein makes it an excellent research tool to visualize the expression of genes and protein distribution in vivo. Furthermore, since the functionality of GFP does not require other proteins or cofactors, and since GFP has not been found to negatively affect other endogenous cell functions, it is functional in a wide variety of both prokaryotic and eukaryotic organisms (Chalfie et al. 1994). *gfp* can be introduced into organisms by integrating the gene into the organism's genome, or by introducing the gene into the organism in an extrachromosomal array.

C. elegans *as a Model Organism*

C. elegans is an excellent model organism to study the regulation of exogenous genes in the germline and soma because expression patterns of fluorescently tagged protein products are easily visualized and imaged under a fluorescent microscope. Furthermore, *C. elegans* have a

relatively fast generation time. Adult animals produce approximately 275 eggs in a 120 hour period (Byerly et al., 1976). Eggs develop into adults in approximately 65 hours (Byerly et al., 1976). Additionally, while *C. elegans* are hermaphrodites, matings can be performed by using male *C. elegans*, which are present at a low frequency in wild-type populations.

Results

Crossing a let-858::gfp *Transgene into* spr-5;met-2 *Double Mutants*

Since Katz et al. had previously shown that when an *h2a::gfp* transgenic array was crossed into *spr-5* mutants, the array was de-silenced in the germline, we asked whether we would see similar de-silencing in *spr-5; met-2* double mutants (Katz et al., 2009). We designed an experiment in which we crossed a *C. elegans* strain with a *let-858::gfp* multi-copy array into an *spr-5; met-2* double mutant background (Figure 1). *let-858* codes for nucampholin, a protein required for *C. elegans* development. *let-858* is expressed in all cell types, however when it is introduced into wild-type animals as an extrachromosomal array it is silenced in the germline (Kelly et al., 1997).

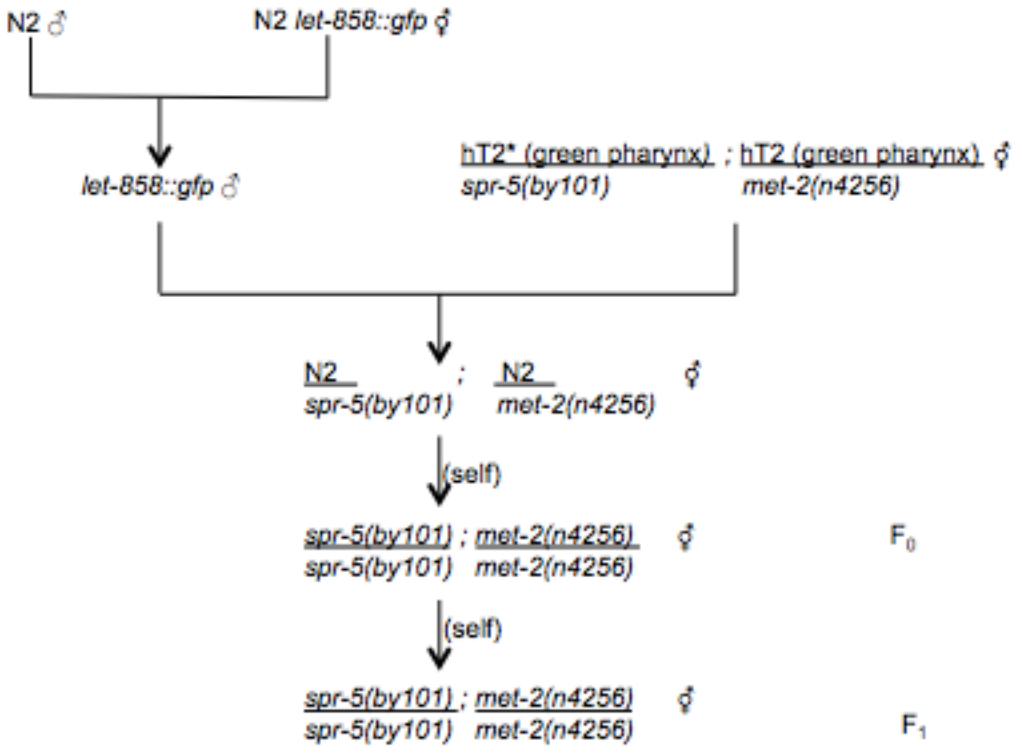

Figure 1: *let-858::gfp x spr-5;met-2* cross design. *Note that the hT2 balancer strain does not completely overlap *spr-5(by101)* on chromosome I, however since the balanced region is genetically close to the spr-5 gene, crossover is less likely.

spr-5;met-2 homozygote double mutants become infertile by the second generation, however double mutant heterozygotes are able to produce viable progeny. We used a balanced *spr-5;met-2* strain (*hT2 [bli-4(e937) let-?(782) qIs48]* strain developed by the Katz lab to maintain the double mutant population as heterozygotes. This balancer strain is a translocation balancer that prevents recombination at locations on both chromosomes I and III. The *met-2* gene is located within the balanced region on chromosome III, however the *spr-5* gene is just outside of the balanced chromosome I region, so loss of the *spr-5* mutation through recombination, is unlikely, though possible. Therefore, the genotype of *met-2 spr-5* double mutants must be either verified via PCR or through observation of the second generation sterility phenotype (Kerr et al., 2014). The first generation of double *spr-5;met-2* mutants are fertile, however 100% of the F_1 progeny will be sterile, and display extreme germline and somatic defects including protruding vulva, multi-vulva, lack of embryo formation. Additionally, the progeny of double mutants are developmentally delayed (Kerr et al., 2014).

We designed the cross as in Figure 1. In the initial stage of the cross we were able to track *spr-5; met-2* heterozygotes by a pharynx GFP marker, in the last stage of the cross we determined if the animals were *spr-5; met-2* heterozygotes or homozygotes by observing the double mutant sterility phenotype. Single *spr-5* and *met-2* mutants occur at a 1/4 ratio in the population, thus we expected to get double mutants at a ratio of 1/16. Therefore, out of 90 plates, we expected to get approximately six double mutants from this cross. From the *let-858* x *spr-5; met-2* cross we found four plates out of 90 that contained double mutants progeny which displayed all of the phenotypes reported by Kerr et al. (Figure 2 G, I) (Kerr et al., 2014). The fact that we found only four double mutants instead of six is not statistically significant ($p=0.414$, chi-square test).

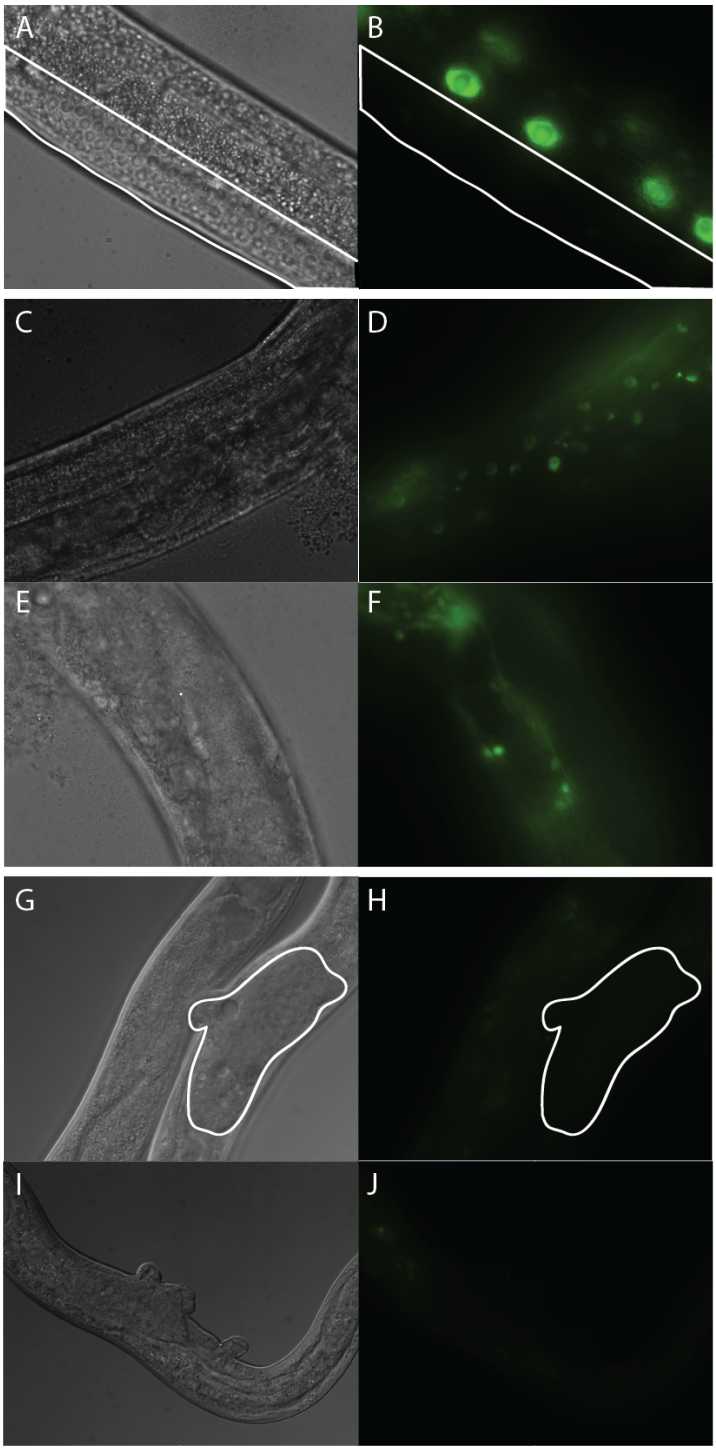

Figure 2: LET-858::GFP expression in wild-type and *spr-5;met-2* F_0 and F_1 homozygous mutant animals. A-B: DIC (A) and FITC (B) image of *let-858::gfp* (*pha-1* (e2123ts) III; ccEx7271). C-F: DIC (C, E) and FITC (D, F) images of *spr-5; met-2 let-858::gfp* double mutant F_0 generation. G-J: DIC (G, I) and FITC (H, J) images of *spr-5; met-2 let-858::gfp* (*pha-1* (e2123ts) III; ccEx7271) double mutant F_1 generation. The germline is enclosed by the white line in A, B, G, and H.

We imaged the F_0 mothers from two of the plates with differential interference contrast microscopy (DIC) and fluorescein isothiocyanate (FITC) microscopy. We found GFP expression in a similar pattern in these F_0 animals as was observed in *let-858::gfp* array-carrying wild-type animals (Figure 2 A-B). The GFP expression in the F_0 generation was weaker compared to the wild-type strain (Figure 2 C-F). However, the GFP expression pattern in the double mutant F_1 progeny was much dimmer than the expression in the F_0 generation, and did not appear to have the nuclear-localized pattern observed in the array carrying wild-type animals (Figure 2 G-J). We found that none of the *spr-5;met-2* double mutant F_1 progeny expressed the *let-858::gfp* transgene in their germline (Figure 2 G-J). Furthermore, all of these F_1 double mutant progeny had little to no visible expression of GFP in somatic cells (Figure 2 G-J, Figure 3). Of the 27 animals that we imaged, only approximately 52% appeared to have any significant GFP signal (Figure 3). We also imaged some of the F_1 progeny with TexasRed (TXRED) fluorescent microscopy (data not shown).

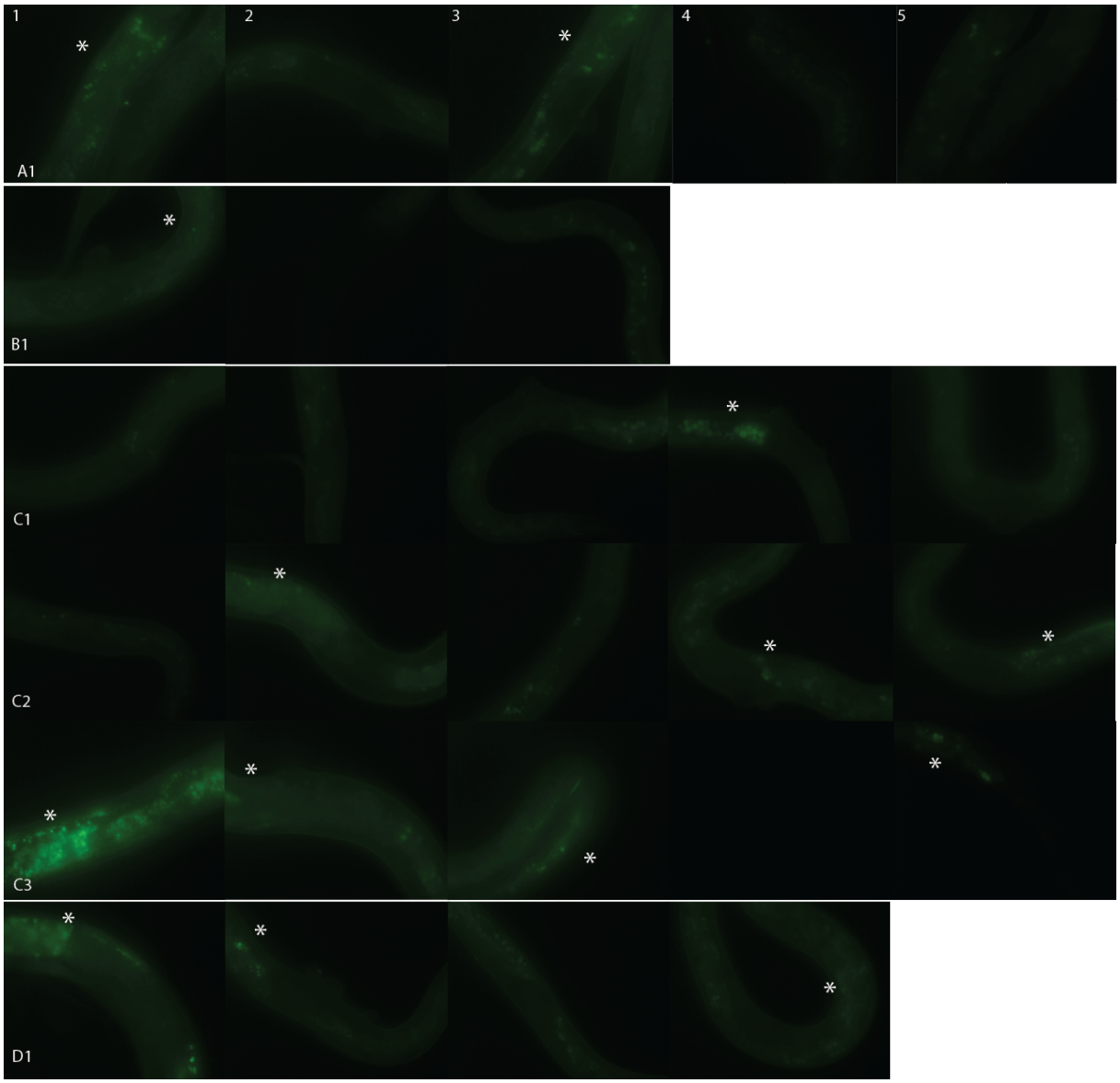

Figure 3: LET-858::GFP Expression in *spr-5;met-2* homozygous mutants. Rows with the same letter indicate F_1 *spr-5;met-2* homozygous animals derived from the same F_0 homozygous *spr-5;met-2* animal. Stars denote visually significant amounts of GFP expression. GFP expression was digitally enhanced equally in all photographs.

To verify that the transgene was still present in the double mutants, we genotyped three animals from each of the four double mutant plates for *gfp*. We found that two animals from two separate plates contained *gfp* DNA (Figure 4). Dr. Teresa Lee repeated the *gfp* genotyping experiment to verify the result, and again found that two animals from two different plates contained the expected GFP DNA band (data not shown). Interestingly, the two animals that had

gfp bands in Dr. Teresa Lee's experiment came from different plates than the animals that were found to have bands in the first genotyping experiment. The significance of this finding will be explored in the results section.

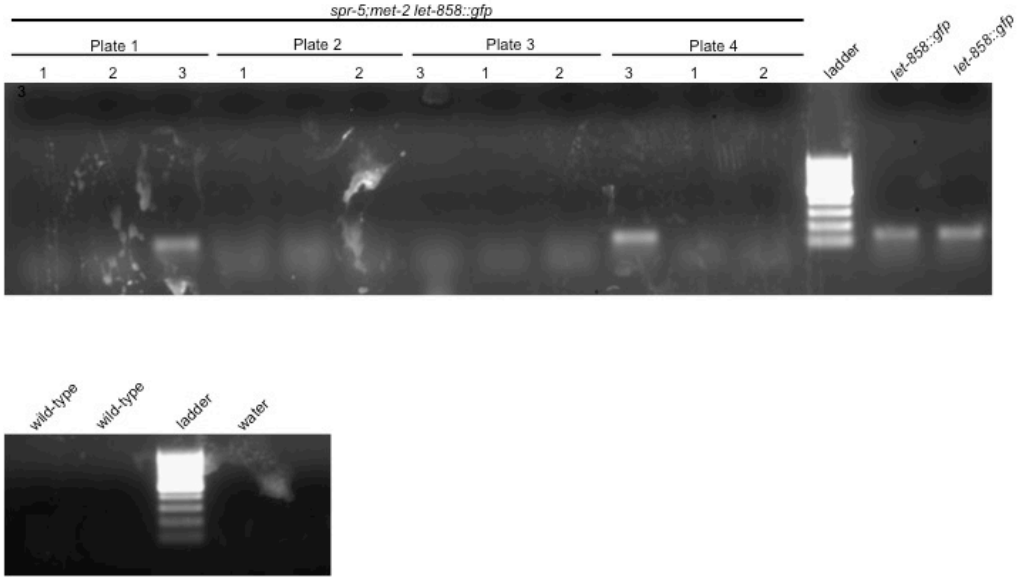

Figure 4: Verification of presence of *let-858::gfp* transgene in *spr-5;met-2* F_0. The DNA from three double mutant F_1 *let-858 spr-5;met-2* animals from each of the four *spr-5;met-2* mutant plates was amplified with *gfp* primers and visualized. The *gfp* band is approximately 150 bp. The genotype of the animal is indicated about the gel lane.

The Biolistically Integrated h2a::gfp *Transgene Maps to Chromosome III*

We also attempted to cross *C. elegans* with an *h2a::gfp* transgene (*unc-119* (ed3) III; *ckls1* (*unc-119* (+) his-53::GFP) into the *spr-5/hT2;met-2/hT2 [bli-4(e937) let-?(782) qIs48]* strain. The *h2a::gfp strain* was obtained from the Kelly lab, and the transgene had been introduced into the *C. elegans* by biolistic transformation. However, we were not certain whether the transgene was integrated into the genome or if it was present as an extrachromosomal array. If the gene were integrated, this cross would potentially provide insight into the relationship between silencing of multi-copy arrays versus integrated transgenes. We carried out this cross in the same way that we did the *let-858::gfp spr-5;met-2* cross. However, out of the 90 somatic GFP expressing F_0 *C. elegans* picked in the last stage of the cross, all animals produced fertile F_1

progeny, indicating that none of the animals were double mutants. This result indicates that the *h2a::gfp* gene was indeed integrated into this strain on either chromosome I or III. Since the double mutant should be present in the population at a 1:16 ratio, we expected to get approximately six double *spr-5; met-2* double mutant animals that produced 100% sterile F_1 progeny from this cross. The fact that we saw no sterile animals is statistically significant (p=0.0177, chi-square test). This result suggests that the *h2a::gfp* gene is indeed integrated into the *C. elegans*, and that it maps to one of the chromosomes on which *spr-5* or *met-2* is located, chromosome I or III. If this is the case, then getting a double mutant which also expresses H2A::GFP would be genetically impossible unless there was a crossover event. Furthermore, using the same *h2a::gfp* strain, Katz et al. were able to cross in *spr-5*, which is located on chromosome I (Katz et al., 2009). Therefore this suggests that the *h2a::gfp* gene construct is integrated on chromosome III.

The Role of spr-5 *in Epigenetic Regulation of Single Copy Integrated Transgenes*

mosSCI is a powerful technique to study both the mechanism of silencing initiation as well as maintenance, since the integrated gene is initially introduced into the animal with no chromatin marks. In the initial paper detailing the mosSCI injection protocol, Frøkjær–Jensen et al. demonstrated that the EG6699 integration strain (which has an insertion site on Chromosome II) allowed for germline expression of integrated transgenes (Frøkjær–Jensen et al., 2008). However the Mello lab reported that germline expression of transgene constructs appeared to be dependent on transgene length, since a *gfp::cdk-1* integrated transgenic construct was silenced in the germline whereas a transgenic construct made with the short *flag* sequence was not expressed in the germline (Shirayama et al., 2012). Further, the Mello lab showed that *prg-1* is implicated

in initiation of epigenetic silencing of the Tc3 transposon family (Shirayama et al., 2012). *prg-1* mutants express single copy transgenes in the germline whereas wild-type animals silence the transgenes in their germline (Shirayama et al., 2012). Since Katz et al. had implicated *spr-5* in the endogenous gene-silencing pathway, we were interested to know whether *spr-5* was involved in the exogenous gene-silencing pathway (Katz et al., 2009). If *spr-5* were involved in non-self gene silencing, we would expect to see a similar de-silencing of a large exogenous gene in the germline of *spr-5* mutants.

Generation of spr-5 mutant mosSCI Injection Strain

We first generated an *spr-5* strain with a mosSCI insertion site by crossing *spr-5 C. elegans* into the EG6699 strain (*unc-119* (III), *ttTI5605* (II)) (crosses performed by Karen Schmeichel). EG6699 *C. elegans* engineered for the mosSCI injection protocol have a *ttTI5605* transposon at the site of integration. Additionally mosSCI *C. elegans* have a mutant copy of *unc-119* necessary for normal movement (Frøkjær-Jensen et al., 2008; Maduro et al., 2000). The injected integration plasmid construct contains *unc-119 (+)*, which serves as a positive selection injection marker, since *C. elegans* that carry a transgene construct either in an extrachromosomal array or integrated into their genome resume wild-type movement (Frøkjær-Jensen et al., 2008).

Karen Schmeichel verified that the mutant *C. elegans* were homozygous for *spr-5*. The verification that the *C. elegans* were homozygous for the mosSCI insertion site was done using primers flanking the transposon insertion site. The *ttTI5605* transposon is approximately 1.3 kb, and animals with the transposon insertion site could be verified by the presence of a 1.3 kb *ttTI5605* PCR product. Multiple plates from the initial cross were tested in order to find a strain that was homozygous for the mosSCI insertion. Eight *C. elegans* from each plate were tested.

Since the 1.3 kb PCR product is produced whenever at least one copy of the transposon is present, plates that had heterozygous populations were expected to yield the 1.3 kb PCR product 75% percent of the time. Plates with homozygous populations were identified molecularly by the presence of a 1.3 kb PCR product in 100% of the fully body lysate DNA samples (Figure 5).

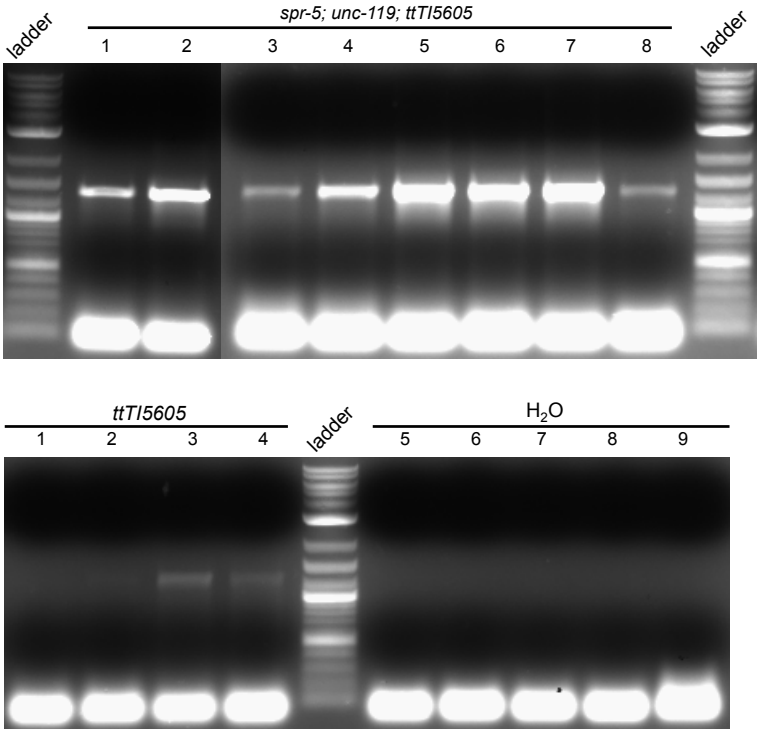

Figure 5: Verification of *ttTI5605* in *spr-5* mosSCI animals. DNA from eight animals was collected and the EG6699 transposon insertion *ttTI5605* was amplified by PCR. All animals from this particular plate showed the expected 1.3 kb PCR product, suggesting that this cross was homozygous for the mosSCI insertion site. Top gel: lanes 1-8 DNA from *spr-5 unc-119 ttTI5605* animals. Bottom gel: lane 1-4 positive control animals known to have *ttTI5605*, lane 5-9 negative water control.

mosSCI Single Copy Insertion in spr-5 *mutants*

In our initial injections of *spr-5* mosSCI *C. elegans* we microinjected with two plasmids: a plasmid containing *gfp*::*cdk-1* and *unc-119* (+), and pCFJ601, which encodes the mos transposase. The *gfp::cdk-1* construct is driven by the endogenous *cdk-1* promoter, which expresses in both somatic and germ cells as CDK-1 is necessary for meiotic and mitotic cell

division (Boxem, 2006). *C. elegans* that were rescued from the uncoordinated phenotype were cloned out and their progeny monitored for heritability of the rescued phenotype. Since extrachromosomal arrays are not transmitted to all progeny, we identified an integration event by looking for lines that consistently produced progeny that were 100% *unc-119* (+).

In order to further verify that the injected *C. elegans* had integrated the exogenous construct and were not maintaining it as an array, we began to perform injections with an additional co-injection marker pCFJ90 (*Pmyo-2::mCherry*), which codes for red fluorescent mCherry expression in the pharynx. *C. elegans* that were maintaining the transgene construct as an array would also maintain the mCherry marker in an array and display pharynx mCherry expression. Integrated *C. elegans* would be expected to express GFP but not display mCherry expression since the array would have been lost.

Out of the approximately 230 unique injections of *spr-5* mutant animals, only five lines were produced which generated *unc-119* (+) progeny (Table 1). We produced four of these lines using the original two-plasmid mixture (no mCherry co-injection marker). These lines appeared to have integrated the transgene based on apparent 100% segregation of *unc-119* (+) to the progeny. The fifth line was created with the mCherry co-injection marker. It is unlikely that the transgene was integrated in this fifth line since all progeny that were *unc-119* (+) also displayed mCherry pharynx expression. The five *unc-119* (+) strains were imaged using DIC, TXRED, and FITC microscopy.

Injection Date	Strain Injected	Plasmid Mixture			Survival/Total Injections	Initial array formation	Maintained array or integration
		Integration Vector	pCFJ601	pCFJ90			
7/29/15	*spr-5*;EG6699	*gfp::cdk-1* – ng/μL	– ng/μL	NI	–/–	Yes	Yes
8/31/14	*spr-5*;EG6699	*gfp::cdk-1* 35 ng/μL	35 ng/μL	NI	–/12	No	No
9/9/14	*spr-5*;EG6699	*gfp::cdk-1* 50 ng/μL	50 ng/μL	NI	–/19	No	No
9/18/14	*spr-5*;EG6699	*gfp::cdk-1* 25 ng/μL	25 ng/μL	NI	–/20	No	No
9/25/14	*spr-5*;EG6699	*gfp::cdk-1* 25 ng/μL	25 ng/μL	NI	-/29	No	No
10/2/14	*spr-5*;EG6699	*gfp::cdk-1* 25 ng/μL	25 ng/μL	NI	6/19	Yes	Yes
10/12/14	*spr-5*;EG6699	*gfp::cdk-1* 25 ng/μL	25 ng/μL	NI	10/13	Yes	Yes
10/17/14	*spr-5*;EG6699	*gfp::cdk-1* 12.5 ng/μL	12.5 ng/μL	NI	9/12	Yes	No
10/24/14	*spr-5*;EG6699	*gfp::cdk-1* 25 ng/μL	25 ng/μL	NI	12/17	Yes	No
10/30/14	*spr-5*;EG6699	*gfp::cdk-1* 10 ng/μL	50 ng/μL	NI	7/11	No	No
11/7/14	*spr-5*;EG6699	*gfp::cdk-1* 10 ng/μL	50 ng/μL	NI	9/17	Yes	No
11/13/14	*spr-5*;EG6699	*gfp::cdk-1* 10 ng/μL	50 ng/μL	NI	8/14	Yes	Yes
11/22/14	*spr-5*;EG6699	*gfp::cdk-1* 10 ng/μL	50 ng/μL	NI	7/17	No	No
3/5/15	*spr-5*;EG6699	*gfp::cdk-1* 50 ng/μL	30 ng/μL	2.5 ng/μL	–/–	Yes	Yes
Unique Injection Days: 14					Total*: 230	8 (3.48%)	5 (2.12%)
10/11/14	*prg-1*; EG6699	*gfp::cdk-1* 25 ng/μL	25 ng/μL	NI	–/13	No	No
10/23/14	*prg-1*; EG6699	*gfp::cdk-1* 25 ng/μL	25 ng/μL	NI	–/15	No	No
11/20/14	*prg-1*; EG6699	*gfp::cdk-1* 10 ng/μL	50 ng/μL	NI	19/22	No	No
Unique Injection Days: 3					Total: 50	0	0
2/3/15	EG6699	*gfp::cdk-1* 50 ng/μL	30 ng/μL	2.5 ng/μL	3/9	No	No
2/10/15	EG6699	*gfp::cdk-1* 25 ng/μL	30 ng/μL	2.5 ng/μL	2/12	No	No
2/12/15	EG6699	*gfp::cdk-1* 50 ng/μL	30 ng/μL	2.5 ng/μL	4/16	No	No
2/17/15	EG6699	*gfp::cdk-1* 25 ng/μL	30 ng/μL	2.5 ng/μL	2/–	No	No
2/19/15	EG6699	*gfp::csr-1* 50 ng/μL	30 ng/μL	2.5 ng/μL	–/–	No	No
2/21/15	EG6699	*gfp::cdk-1 50* ng/μL	30 ng/μL	2.5 ng/μL	8/21	No	No
Unique Injection Days: 6					Total*: 88	0	0
Unique Injection Days: 23					Overall Total* 350	8	5

Table 1: mosSCI injections. Some data for plasmid mixture concentration and number of injected and surviving animals is incomplete, indicated by –. *Note: To calculate the total unique injections for strains with incomplete data, 15 animals were assumed to have been injected on these days. As a result, the unique injection total for strains with incomplete data is not exact.

None of the animals in the five strains that had the *gfp::cdk-1* transgene (either integrated or as an extrachromosomal array) displayed germline GFP expression (Figure 6A-E). Of those that did express GFP in their somatic cells, GFP was often brightest in the tail and pharynx regions (Figure 6F), as well as in some of the later stage embryos (both retained and laid) (Figure 6G). Furthermore, the GFP that was expressed in the somatic cells of *spr-5* mutants was relatively weak, and some animals did not appear to express GFP at all. Transgenic lines had to be maintained by picking *C. elegans* that had the brightest GFP expression. Interestingly, in the strain that was obtained by co-injecting with the mCherry marker, pharynx expressing mCherry marker was very bright in all *unc-119* (+) *C. elegans*, even when GFP expression was low or nonexistent (Figure 7).

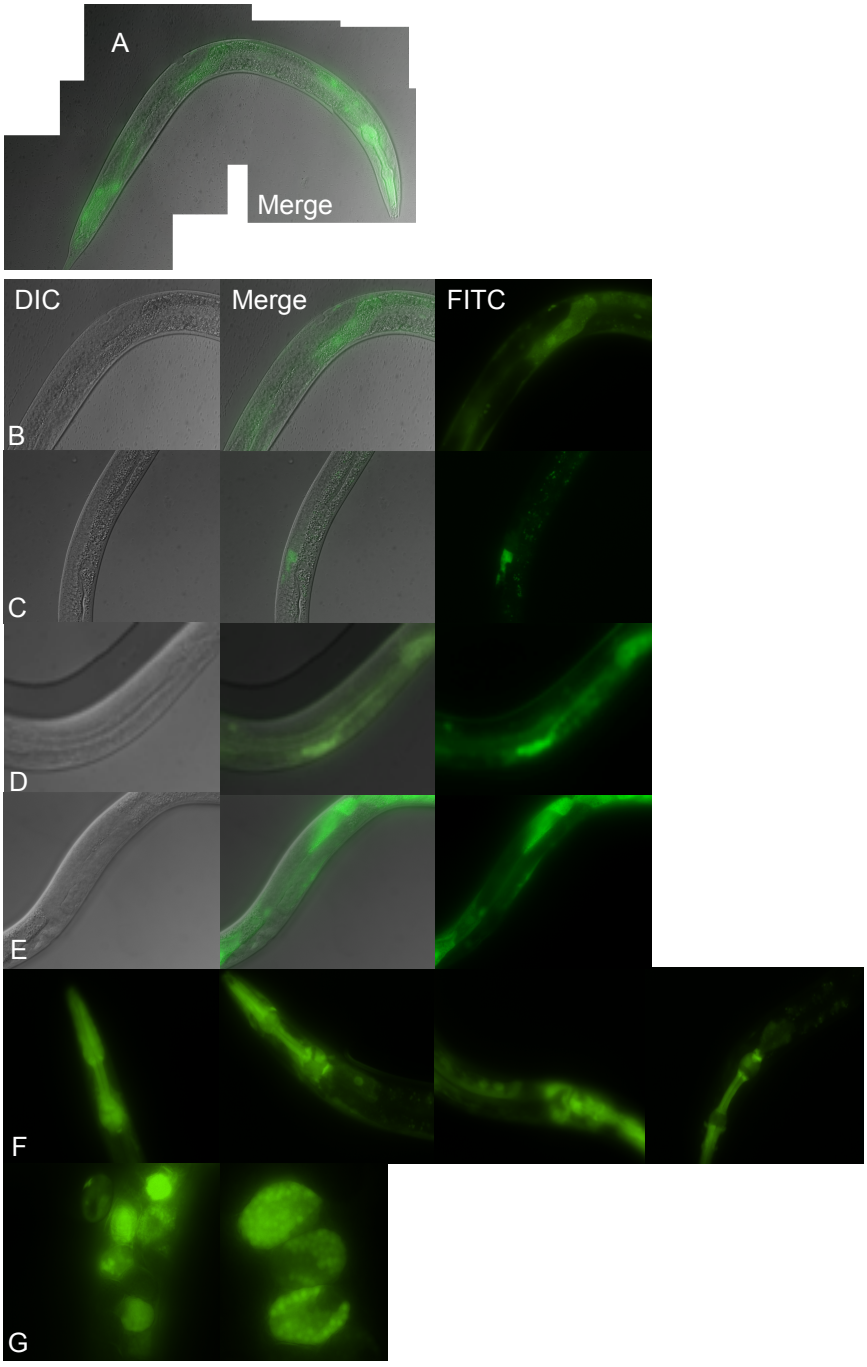

Figure 6: GFP expression in *spr-5* animals injected with *gfp::cdk-1* A. DIC, Merge (of DIC and FITC) and FITC of *spr-5* animal from 10/12 *gfp::cdk-1* injection line, B-E. Germline of *spr-5* animal from B. 7/29, C. 10/2, D. 10/12, E. 11/13 *gfp::cdk-1* injection lines. Expression observed in somatic cells but not germline. F. Strong GFP expression in the pharynx of *spr-5 gfp::cdk-1* injected animals (from left to right, injection strains 7/29, 10/2, 10/12, 11/13). G. GFP expression in *spr-5* mutant embryos. Note that FITC images were digitally enhanced to improve GFP visualization.

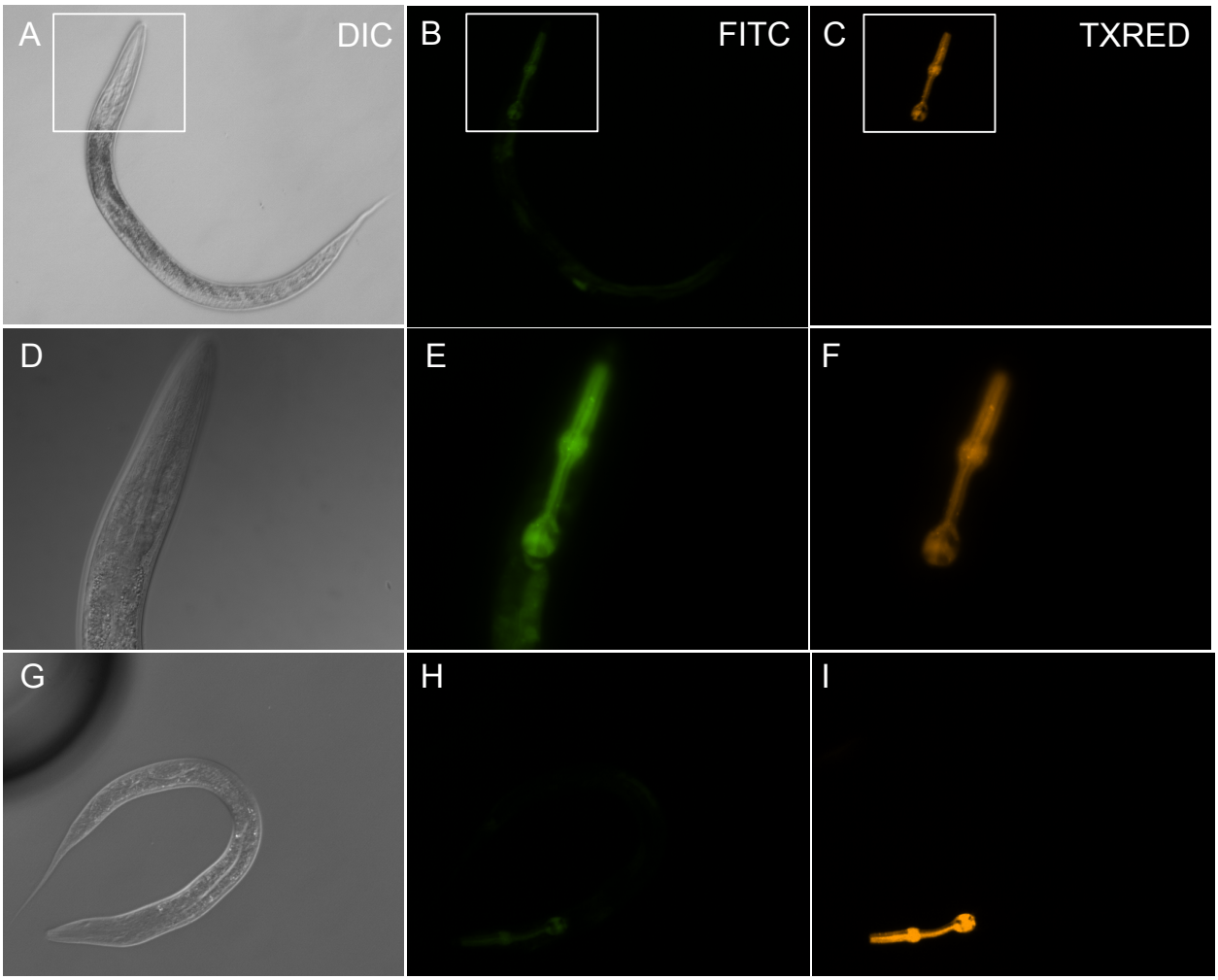

Figure 7: GFP and mCherry in *spr-5* mosSCI strain injected with *gfp::cdk-1*, pCFJ601 and mCherry pCFJ90 co-injection marker. A-C: Full body expression pattern A. DIC, B. FITC C. TXRED. D-E: Magnified image of highlighted pharynx region in A-C D. DIC, E. FITC, F. TXRED. G-I: Full body expression patterns of GFP and mCherry G. DIC, H. FITC, I. TXRED.

As a positive control experiment, we injected *gfp::cdk-1*, pCFJ601, and pCFJ90 into EG6699 mosSCI *C. elegans* with no additional background mutations (*unc-119* (III), *ttTI5605* (II)). The EG6699 line was obtained from the Benian lab. All animals displayed the *unc-119* phenotype, but we did not verify the presence of the *ttTi5605* transposon. This injection was expected to produce animals which expressed the *gfp:cdk-1* construct somatically but silenced it in the germline. None of the *unc-119* (III), *ttTI5605* (II) injected *C. elegans* were found to be rescued from the *unc-119* phenotype via array formation or integration.

Discussion

Apparent Silencing of let-858::gfp *in* spr-5;met-2 *Homozygous Mutants*

The lack of *let-858::gfp* expression in the F_1 population of *spr-5;met-2* mutants from the cross experiment could be explained two ways: (1) the *let-858::gfp* extrachromosomal array was lost in the F_1 progeny or (2) the *let-858::gfp* extrachromosomal array was silenced in the F_1 progeny due to the *spr;5;met-2* double mutation.

In our cross experiment of the *let-858::gfp* transgene into *spr-5;met-2* double mutants, while the GFP expression was indeed consistently faint (Figure 3), we do believe that this was completely due to loss of the *let-858::gfp* transgene. Firstly, after enhancing the FITC images of the *spr-5;met-2* double mutants so that the GFP expression could be better visualized, the GFP expression appeared to be more localized to specific features within the F_1 *C. elegans* double mutants than the diffuse expression pattern typical of auto-fluorescence (Figure 3 row A1 column 1, C3x1, D1x1), suggesting that the fluorescence was true GFP. Secondly, when most of the *C. elegans* were imaged using TXRED, little to no fluorescence signal was observed. This result indicates that the small amount of fluorescence that was observed in some of the animals was not completely due to auto-fluorescence since auto-fluorescence is often apparent under the TXRED lens, while true GFP expression is not. Thirdly, we observed that both F_0 mothers had *gfp* expression. The reason that GFP expression was fainter in the F_0 animals (Figure 2 D, F) than normal *let-858::gfp* animals (Figure 2 H, J) may have been to the F_0 parent's advanced age when it was imaged. Fourthly, we molecularly verified that the GFP gene was present in at least some of the animals on each plate (Figure 4 and unpublished data from Dr. Teresa Lee). These results indicate that the fluorescence that we observed in the *let-858 spr-5;met-2* mutants was due to actual GFP expression, that at least some of the double mutants on each of the four plates still

contained the *let-858::gfp* plasmid. Combined, these findings indicate that LET-858::GFP expression was silenced somatically in some of the *spr-5;met-2* double mutant animals.

While our molecular verification of the presence of GFP in the double mutant populations did confirm the presence of GFP in at least some of the animals in each of the four double mutant populations, it was puzzling that we only saw GFP bands in approximately 16.7% of the C. elegans that we genotyped for *gfp* (Figure 4, data from Dr. Teresa Lee not shown). This low observed percent of double *spr-5; met-2* mutant C. elegans with GFP bands could be a result of low extrachromosomal array transmission or could be a result of errors in the PCR reaction. When maintaining the wild-type *let-858::gfp* population, we observed a seemingly high rate of extrachromosomal array segregation. However, as this strain carries a *pha-1* mutation which is lethal but rescued for by the *let-858::gfp* construct, the apparent construct inheritance rate is skewed since animals that do not inherit the construct die. Therefore in order to verify the true array inheritance rate of the *let-858::gfp* array, we would need to analyze the array transmission rate in population of C. elegans carrying the *let-858::gfp* array without the lethal mutation. If, after this analysis, we found that the double *spr-5; met-2* mutants still have a lower array transmission rate than wild-type animals, this could indicate that in addition to a role in sustaining a more active chromatin state, either *spr-5* and/or *met-2* is implicated in the propagation of extrachromosomal arrays. The low rate of animals with *gfp* bands could also result from the reduced amount of DNA in double mutants, which have significantly fewer cells due to their reduced germline size phenotype. Furthermore, this genotyping result could have been due to issues with the *gfp* primers, however this is unlikely since both positive controls showed strong *gfp* bands (Figure 4).

Lack of Formation of Arrays in spr-5 *mosSCI Injections*

Our mosSCI injections efficiency was unusually low. The Jorgenson lab reported that each injected animal typically produced about 10-15 array-carrying, *unc-119* (+) animals (Frøkjær-Jensen WormBuilder website, 2013). In our hands, however, essentially the only rescued animals that we produced were those in the five transgenic lines. Our low array formation could be explained two ways: (1) our injection protocol and/or technique was flawed or (2) the *spr-5* mutation interfered with the animals' ability to form transgenic arrays because it led to excessive silencing of the transgene in both the germline and soma.

It is quite possible that there was something wrong with our injection protocol. For example, the plasmid mixture used in the injections could have also contained a plasmid at a toxic concentration, killing *C. elegans* that successfully recieved the array. Additionally, the plasmid concentrations could have been not amenable to integration. To address these concentration-based issues, we tried optimizing the plasmid mixture concentrations, but were unable to increase the injection efficiency. Furthermore, it is possible that there were problems with my injection technique. However, it is unlikely that this was the source of problem since multiple people observed my injection technique and verified that I was injecting into the correct part of the gonad. Nevertheless, we cannot rule out the possibility that our flawed technique was the cause of this low rate, particularly since we were unable to successfully form arrays in our *gfp::cdk-1* into EG6699 positive control injections. To continue to trouble shoot this procedure for protocol issues, we should verify that the *ttTI5605* (II) transposon is present in the control EG6699 animals (no *spr-5* mutation), and continue to inject into these animals until we are able to get normal, widespread array formation.

spr-5;met-2 *Mutants Silence Non-Self DNA Somatically*

The second possible explanation for both our results from both the mating experiment and mosSCI injections is that either *spr-5* and/or *met-2* is implicated in differentiating between germline and somatic identity, thus impacting whether a transgene is silenced or not in these two different environments. The presence of the *let-858::gfp* array in at least some of the F_1 *spr-5;met-2* double mutants and yet the lack of robust GFP expression, combined with the low array formation in the mosSCI injection experiments suggests a role for *spr-5* and *met-2* in maintaining an active chromatin state in wild-type animals.

If *spr-5* is implicated in the silencing pathway of transgenes, this could explain our inability to form *spr-5* mosSCI integrants. If, even when animals were successfully injected with the array, they immediately silenced the transgene construct and other co-injection plasmids, an integration event would be highly unlikely. While we did obtain five strains that appeared to express and propagate the transgene, it is possible that these resulted from a leaky *spr-5* phenotype. Additionally, even when an array or integration was achieved, GFP expression was not robust, and not uniform throughout the body, suggesting that some somatic silencing was occurring (Figure 6 A-E). Indeed, as discussed earlier, a GFP-expressing population was only maintained by cloning out the brightest animals every generation.

Additionally, our results from the mating experiment support the model that *spr-5* and/or *met-2* are maternally deposited, and thus the active transcription state of transgenes in somatic cells facilitated by *spr-5* and/or *met-2* is dependent on passage through the germline (Kerr et al., 2014). We observed that the F_0 double *spr-5;met-2* mutant generation did not have significant silencing of *gfp*, consistent with the idea that even though these animals did not have functional copies of *spr-5* or *met-2*, they were able to still properly express the transgene somatically

because of the inheritance of functional *spr-5* and *met-2* transcripts from their heterozygous parents (Figure 2 C-F). However, the F_1 population showed extensive silencing of the transgene. This silencing could be explained by the fact that these F_1 animals neither had functional copies of *spr-5* or *met-2*, nor inherited functional mRNA transcripts of these genes maternally (Figure 2, G-J).

Our unexpected results from the mating experiment, combined with a new interpretation of our results from the mosSCI injection experiments, caused us to change our focus from regulation of exogenous genes specifically in the germline, to the mechanism through which cells distinguish between somatic and germline identity and regulate exogenous genes in the soma versus in the germline. We propose three models, one which implicates *spr-5* and/or *met-2* in a direct role in transgene silencing fate in the soma and germline, and two which implicate these genes in an indirect role in silencing fate in the soma and germline.

Direct Action Model

spr-5 and/or *met-2* could work directly in somatic tissues to make a more active chromatin state in this cell population. Previous work has shown that both *spr-5* and *met-2* are only present in significant amounts in the germline, which makes it difficult to imagine how these proteins are interacting with chromatin in the soma (Kerr et al., 2014). Nevertheless, it is possible that *spr-5* and/or *met-2* are present in the soma in a low but still functional concentration, rendering them undetectable in previous studies. However, this direct model is further complicated by the fact that both enzymes are involved in creating repressive chromatin state in the germline, whereas we observed a more repressive chromatin state in mutants. Therefore, this direct action model would require that the function of both *spr-5* and *met-2* be

essentially reversed in the soma. Therefore, it is unlikely that these enzymes affect the silencing state of exogenous DNA directly in somatic cells.

Indirect Action Model 1: met-2 *and* spr-5 *are Implicated in the RNAe Pathway*

Another model for the action of *met-2* and *spr-5* is that they somehow interact with the RNAe pathway (Shirayama et al., 2012). The current understanding of the RNAe pathway is that *C. elegans* could differentiate between self and non-self DNA in the germline by passing on a "library" of self-RNA transcripts from generation to generation. This library is then compared to expressed transcripts, and used to target and shut down foreign elements such as transgenes and transposons. *met-2* and/or *spr-5* may function either directly or indirectly in the proper inheritance of a functional RNAe reference library from the mother in the germline. Double mutants may improperly utilize the germline RNAe library in their somatic cells in addition to their germline, thus leading to silencing of transgenic elements in the soma which are usually only silenced in the germline. This model is supported by our observation that *let-858::gfp* expression in the double mutant mother in a similar expression pattern as wild-type *let-858::gfp* array-carrying animals (Figure 2 A-F), whereas the double mutant's F_1 progeny has a decidedly different and more silenced GFP expression pattern (Figure 2, G-J). This suggests that the silencing effect may be dependent on passage through the germline, which could implicate *met-2* and/*or spr-5* in the RNAe pathway.

Indirect Action Model 2: spr-5 *and* met-2 *are Necessary to Silence Germline Specific Transgene Silencer Gene(s) and to Determine Somatic vs. Germline Transcriptional Patterns*

Another way in which *spr-5* and *met-2* could work together to return to the pluripotent epigenetic state of the fertilized embryo is through an interaction with another, yet unidentified, "Gene X" which is implicated in germline silencing of transgenes (Figure 8). Previous work by the Katz lab has shown that *spr-5* and *met-2* are required to reset the epigenetic state of spermatogenesis genes between generations. Therefore, it is likely that these proteins also work to silence other germline specific genes such as "Gene X". In our model, the germline of wild-type animals initially has active H3K4me marks. One of the genes which is activated by these marks is a germline expressed Gene X implicated in the silencing pathway of exogenous DNA. These active H3K4 marks are removed by *spr-5* and repressive H3K9me marks are added by *met-2* upon fertilization (passage through the germline). This repressive chromatin state prohibits expression of Gene X, and, as a result, the un-repressed transgene is expressed in somatic cells. As cells differentiate into somatic and germline cells, H3K4me is put back onto genes such as Gene X in the germline to activate germline silencing of the transgene, whereas the repressive marks put on by *spr-5* and *met-2* propagate in the somatic cells, causing them to continue to silence Gene X and thus to continue to express the transgene.

In *spr-5;met-2* double mutants however, H3K4me is not removed, and H3K9me is not added upon fertilization. As a result, Gene X and germline specific genes continue to express in the fertilized egg. This methylation pattern is propagated in somatic cells, causing somatic genes to improperly express Gene X, which causes silencing of the transgene somatically (as well as in the germline). These somatic cells would be expected to show misregulation of germline specific genes and potentially somatic genes as well. This model is reminiscent of the interaction between

spr-5, *hop-1* and *sel-12*. *spr-5* has been shown to repress the somatically expressed *hop-1* gene. When expressed, HOP-1 represses *sel-12*. In *spr-5* mutants, however, *hop-1* is not repressed, thus *sel-12* is ectopically silenced, much in the same way that we saw ectopic silencing of the transgene in somatic cells of *spr-5* mutants (Eimer et al., 2002).

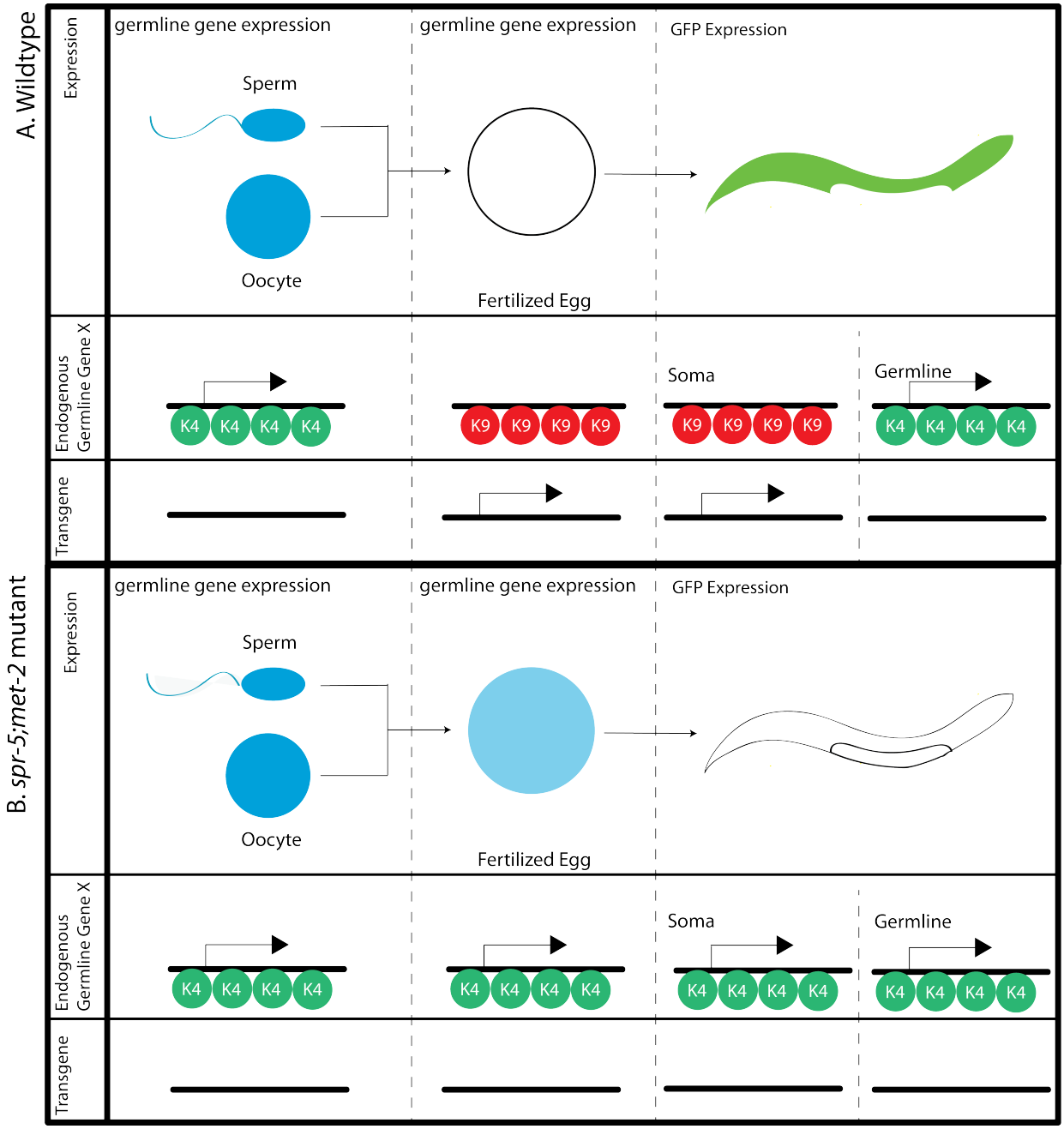

Figure 8: Model of *spr-5* and *met-2* in regulation of silencing in germline and somatic cells. Green K4 circles represent H3K4 methylation. Red K9 circles represent H3K9 methylation. A. Wild-type animals B. *spr-5; met-2* double mutants

Future Work

In addition to misregulation of transgenes, somatic genes may be improperly regulated in the soma of double *spr-5;met-2* mutant animals because of the improperly active state of certain germline specific transcriptional regulation genes promoting germline-like transcription patterns in the soma. To investigate this, we could perform a quantitative PCR (qPCR) experiment to determine the changes in expression patterns of a select group of somatic genes in somatic cells of double *spr-5;met-2* mutant somatic cells. If this qPCR experiment suggested that somatic gene expression patterns were significantly decreased in somatic cells, we could then perform an RNA sequencing experiment using somatic cells from *spr-5;met-2* double mutants to determine how expression levels of both somatic and germline genes are affected.

More work should also be done to distinguish between the roles of *met-2* and *spr-5* in the silencing pathway of exogenous genes. It would be interesting to determine whether *spr-5* and/or *met-2* are involved in initiation and/or maintenance of silencing, as these pathways are distinct. For example, it is possible that SPR-5 suppresses a gene implicated in the initiation of silencing of exogenous DNA in the germline since *spr-5* mutants were able to initiate silencing of the injected transgene in the germline and sometimes partially in the soma in our mosSCI injection experiments of single mutants. Additionally, *spr-5;met-2* double mutants were able to initiate silencing somatically, further implicating *spr-5* and/or *met-2* in a silencing initiation regulation pathway.

In order to further understand the roles of *spr-5* and *met-2* in the transgene silencing pathway, and the role of maternal deposition of functional mRNA transcripts of these genes, we could repeat the *let-858::gfp spr-5;met-2* double mutant cross experiment, but cross out the F_0 generation of *spr-5;met-2* mutants with wild-type males. If the F_1 *spr-5;met-2* heterozygous

progeny from this cross do not express the *let-858::gfp* transgene somatically, displaying a similar somatic silencing phenotype as observed in F_1 double mutants, this would indicate that the action of *spr-5;met-2* is dependent on maternal deposition of mRNA transcripts of these genes.

It would also be worthwhile to optimize the mosSCI injection procedure so that we can get successful injections into the EG6699 strain (no mutant background). By creating this control, we could better compare the expression levels of *gfp::cdk-1* in *spr-5* mutants to the baseline expression levels of *gfp::cdk-1* in EG6699 *C. elegans*. Furthermore, we should verify molecularly whether *gfp::cdk-1* is present in these animals as an extrachromosomal array, or as an integrated single copy gene, since the silencing mechanisms for extrachromosomal arrays and integrated single copy genes are potentially distinct.

Additionally, work should be done to more fully understand the silencing that we observed in the *spr-5;met-2* double mutants. In order to definitively show that the lack of GFP signal in the double mutants was indeed due to silencing of the DNA rather than loss of the array, we would like to replicate the cross of *let-858::gfp* into the double mutant background and image the F_0 mother and F_1 progeny generations using the same exposure length in order to normalize the GFP expression. Using these normalized images we will be able to quantitatively analyze the expression data. Furthermore, we would like to take age matched images of the F_0 double mutant mothers and genotype these mothers for the presence of GFP. This will allow us to determine the genetic presence of GFP in these animals, and further to quantitatively analyze GFP expression levels.

We would also like to look further into why the strongest GFP expression in *spr-5* single mutants was observed in the pharynx and tail, and why mCherry expression was so consistently

strong in the pharynx. It may be that the mechanism that differentiates between soma and germline also is implicated in distinguishing between different somatic cell populations. Indeed, these expression patterns may suggest that the pathway for somatic transgene expression is different in somatic pharynx cells than in non-pharynx somatic cells.

Previously SPR-5 and MET-2 were implicated in the formation of a more repressive endogenous chromatin environment, yet here we find that mutants in these genes experience repression of exogenous DNA (Kerr et al., 2014). This finding indicates that the methylation remodeling carried out by SPR-5 and/or MET-2 is part of a complex pathway that affects genes that are involved in the regulatory pathway of exogenous DNA. Furthermore, as the action of these two proteins appears to be specific to the germline, it seems that SPR-5 and MET-2 work together to distinguish between germline and somatic transcriptional patterns, thus determining the transcriptional response to the introduction of exogenous genes (Kerr et al., 2014). These findings implicate SPR-5 and MET-2 in the regulation of exogenous DNA in addition to the role that these proteins play in endogenous germline genes regulation. Furthermore, our results suggest that SPR-5 and MET-2 may be implicated in determination of the chromatin environment in germline and somatic cells.

Experimental Procedures

C. elegans Maintenance

C. elegans strains were maintained on OP50 E. coli bacteria on 6 cm nematode growth media plates at 20˚C.

PCR Genotyping Verification

spr-5 mutants were verified using F: 5' - AAA CAC GTG GCTCCA TGA AT - 3' R: 5' - GAG GTT TTG AGG GGT TCC AT - 3'. The *ttTI5605* transposon insertion in the EG6699 strain was verified using F: 5' – CAT GAT GGT AGC AAA CTC ACT TCG – 3' R: CGA AAC AAACAC TGA TCTTAC TTG C – 3'. The presence of GFP was verified with primers generously provided by the Kelly lab.

Injection plasmid preparation

gfp:cdk-1, *gfp::csr-1*, pCFJ601 and pCFJ90 were prepared using a Qiagen QIAprep Spin Miniprep kit.

gfp::cdk-1 plasmid

Shirayama et al. generated a plasmid which contained a single copy of *cdk-1* annealed to *gfp* by creating an restriction digest site by site-directed mutagenesis prior to the *cdk-1* transcription start site, and inserting *gfp* sequence there in a pCFJ151 vector (*gfp::cdk-1*). We obtained this plasmid from the Mello lab and verified it by restriction enzyme digest.

Creation of *spr-5* mosSCI Strain

Dr. Karen Schmeichel developed the *spr-5* mosSCI strain by crossing *spr-5 (by101)* mutants into the EG6699 mosSCI background. The *spr-5 (by101)* strain was originally obtained from R. Baumeister. *spr-5* animals produce a mutant protein because there is a transposable element inserted in a coding region of the gene at 13447 kbp.

mosSCI injection procedure

The mosSCI microinjection protocol developed by the Jorgenson lab allows for a targeted single copy insertion of the gene of interest into *C. elegans* by taking advantage of a targeted transposon insertion site. mosSCI strains have been engineered to have a *ttTI5605* transposon located at a specific genetically neutral site in the genome that is excised when a plasmid containing a gene coding for transposase is introduced into the animal. The double strand break is then repaired with a co-injected plasmid containing the gene of interest flanked by a sequence which is recognized by the double break site. The Jorgenson lab has developed *C. elegans* strains with multiple mosSCI integration locations, however previous work has shown the strain with the chromosome II integration site to have the highest rate, 43%, of integration compared to the other integration locations (Frøkjær-Jensen et al., 2012). Approximately 15-20 hours before the microinjection was performed, non-starved, healthy looking *C. elegans* in larval stage 4 (L4) were transferred to a plate of OP50 E. coli and kept at 16 ˚C in order to increase injection survival rate. The injections were done with plasmid mixtures of varying concentrations (Table 1). After being injected, surviving animals were placed on a fresh P50 plate and allowed to recover at 16 ˚C for 1-10 hours. Animals that were found to be still alive after this time period were transferred to their own plate. After approximately four days the plates were screened for animals with *unc-119* (+) movement. Plates were screened up to approximately five days after the plates were starved out. F_0, the injected animal, potentially had oocytes where the transgene has integrated or formed an array. If an integration or array formation event occurred, the F_1 progeny would be a mixture of normal movement and mutant movement. Those animals with normal movement would either be heterozygous for the mutation or carry an extrachromosomal array. The F2 progeny of these heterozygous *C. elegans* will be a mixture of *unc-119* (+/+), *unc-*

119 (-/-) and *unc-119* (+/-) animals. *unc-119* (+/?) F2 *C. elegans* from these plates were then cloned out, and the F3 progeny were monitored. F3 clones heterozygote for the insertion gave rise to an F4 progeny mixture of *unc-119* (+/+), *unc-119* (-/-) and *unc-119* (+/-), whereas F3 clones homozygote for the insertion gave rise to 100% *unc-119* (+/+) progeny. Array carrying animals however were expected to give rise to a population that was less than 100% unc-119 (+/+) since extrachromosomal arrays do not segregate perfectly. Animals that were thought to be homozygote were maintained over multiple weeks, during which time the segregation of unc-119 (+/+) was monitored.

Genetic Crosses

The KW1128 *let-858::gfp* strain (*pha-1* (e2123ts) III; ccEx7271) and KW1044 *h2a::gfp* (*unc-119* (ed3) III; *ckls1* (*unc-119* (+) his-53::GFP) transgenic strains were obtained from the Kelly lab. ccEx7271 is an extrachromosomal array composed of many repeats of *let-858::gfp* (pBK48.1). These transgenes were crossed into wild-type Bristol N2 males obtained originally from the *Caenorhabditis* Genetic Center. Transgene carrying males were mated with a double mutant *spr-5(by101);met-2(n4256); hT2 [bli-4(e937) let-?(782) qIs48]* balanced strain maintained as a heterozygote population. *C. elegans* were imaged using DIC and, FITC, and TXRED fluorescent microscopy.

Sources

Andersen, E.C., and Horvitz R (2007). Two C. elegans histone methyltransferases repress lin-3 EGF transcription to inhibit vulval development. Development *134*, 2991-2999.

Bannister, A.J., Schneider R, and Kouzarides T (2002). Histone methylation: dynamic or static?. Cell, *109*, 801-806.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

Bessler, J. B., Andersen, E. C., and Villeneuve, A. M. (2010). Differential localization and independent acquisition of the H3K9me2 and H3K9me3 chromatin modifications in the Caenorhabditis elegans adult germ line. PLoS Genetics *6*, e1000830.

Boxem, M. (2006). Cyclin-Dependent Kinases in *C. Elegans*. Cell Division *1*, 6.

Byerly, L., Cassada, R.C., and Russell, R.L. (1976) The life cycle of the nematode Caenorhabditis elegans: I. Wild-type growth and reproduction. Developmental Biology 51, 23-33.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.

Das, P.P., Bagijn, M.P., Goldstein, L.D., Woolford, J.R., Lehrbach, N.J., Sapetschnig, A., Buhecha, H.R., Gilchrist, M.J., Howe, K.L., Stark, R., Matthews, N., Berezikov, E., Ketting, R.F., Tavaré, S., and Miska, E.A. (2008). Piwi and piRNAs Act Upstream of an Endogenous siRNA Pathway to Suppress Tc3 Transposon Mobility in the *Caenorhabditis elegans* Germline. Molecular Cell *31*, 79-90.

Dillon, S.C., Zhang, X., Trievel, R.C., and Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biology *6*, 227.

Eimer, S., Lakowski, B., Donhauser, R., and Baumeister, R. (2002). Loss of spr-5 bypasses the requirement for the C. elegans presenilin sel-12 by derepressing hop-1. The EMBO Journal *21*, 5787-5796.

Frøkjær-Jensen, C. (2013). Comments/FAQ. WormBuilder: Modifying the C. Elegans Genome. Retrieved from http://www.wormbuilder.org/test-page/commentsfaq/

Frøkjær-Jensen, C., Davis, M.W., Ailion, M., and Jorgensen, E.M. (2012). Improved Mos1-mediated transgenesis in C. elegans. Nature Methods *9*, 117-118.

Frøkjær-Jensen, C., Davis, M.W., Hopkins CE, Newman, B.J., Thummel, J.M., Olesen, S.P., Grunnet, M., and Jorgensen, E.M. (2008). Single-copy insertion of transgenes in Caenorhabditis elegans. Nature Genetics *40*,1375-1383.

Hightower, L.E. (1991). Heat shock, stress proteins, chaperones, and proteotoxicity. *Cell 66*, 191-197.

Holoch, D., and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. Nature Reviews Genetics *16*, 71-84.

Kaltwasser, M., Wiegert, T., and Schumann, W. (2002). Construction and application of epitope- and green fluorescent protein-tagging integration vectors for Bacillus subtilis. Applied and Environmental Microbiology *68*, 2624-2628.

Katz, D.J., Edwards, T.M., Reinke, V., and Kelly, W.G. (2009.) A *C. elegans* LSD1 Demethylase Contributes to Germline Immortality by Reprogramming Epigenetic Memory. Cell *137*, 308-320.

Kelly, W.G., and Fire, A. (1998). Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans. Development *125*, 2451-2456.

Kelly, W.G., Xu, S., Montgomery, M.K., and Fire, A. (1997). Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. Genetics *146*, 227-238.

Kelly, W.K., O'Connor, O.A., and Marks, P.A. (2002). Histone deacetylase inhibitors: from target to clinical trials. *Expert Opinion on Investigational Drugs* 11, 1695-1713.

Kerr, S.C., Ruppersburg, C.C., Francis, J.W., and Katz, D.J. (2014). SPR-5 and MET-2 function cooperatively to reestablish an epigenetic ground state during passage through the germ line. Proceedings of the National Academy of Sciences *111*, 9509-9514.

Ketting, R.F., Haverkamp, T.H., van Luenen, H.G., and Plasterk, R.H. (1999). *mut-7* of *C. elegans*, Required for Transposon Silencing and RNA Interference, Is a Homolog of Werner Syndrome Helicase and RNaseD. Cell *99*, 133-141.

Kornberg, R.D. (1974). Chromatin structure: a repeating unit of histones and DNA. Science *184*, 868-871.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693-705.

Kutter, C., and Svoboda, P. (2008). miRNA, siRNA, piRNA: Knowns of the unknown. RNA Biology *5*, 181-188.

Lakowski, B., Roelens, I., and Jacob, S. (2006). CoREST-like complexes regulate chromatin modification and neuronal gene expression. Journal of Molecular Neuroscience *29*, 227-239.

Lee, H.C., Gu, W., Shirayama, M., Youngman, E., Conte, D., and Mello, C.C. (2012). *C. elegans* piRNAs Mediate the Genome-wide Surveillance of Germline Transcripts. Cell *150*, 78-87.

Lee, M.G., Wynder, C., Cooch, N., and Shiekhatter, R. (2005). An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. Nature *437*, 432-435.

Maduro, M.F., Gordon, M., Jacobs, R., and Pilgrim, D.B. (2000). The UNC-119 family of neural proteins is functionally conserved between humans, Drosophila and C. elegans. Journal of Neurogenetics *13*, 191-212.

Morris, K.V., Chan, S.W.L., Jacobsen, S.E., and Looney, D.J. (2004). Small interfering RNA-induced transcriptional gene silencing in human cells. Science *305*, 1289-1292.

Nottke, A.C., Beese-Sims, S.E., Pantalena, L.F., Reinke, V., Shi, Y., and Colaiácovo, M.P. (2011). SPR-5 is a histone H3K4 demethylase with a role in meiotic double-strand break repair. Proceedings of the National Academy of Sciences *108*, 12805-12810.

Poulin, G., Dong, Y., Fraser, A.G., Hopper, N.A., and Ahringer, J. (2005). Chromatin regulation and sumoylation in the inhibition of Ras-induced vulval development in Caenorhabditis elegans. The EMBO Journal *24*, 2613-2623.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953

Shirayama, M., Seth, M., Lee, H.C., Gu, W., Ishidate, T., Conte, D., and Mello, C.C. (2012). piRNAs initiate an epigenetic memory of nonself RNA in the C. elegans germline. Cell *150*, 65-77.

Vakoc, C.R., Mandat, S.A., Olenchock, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin. Molecular Cell *19*, 381-391.